

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 28, 2003, 18:27:17 ; Search time 33.7273 Seconds

(without alignments)
107.116 Million cell updates/sec

Title: US-09-743-225-9

Perfect score: 73

Sequence: 1 KDKATFGTHDGGXA 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:**

- 1: sp.archaea:**
- 2: sp.bacteria:**
- 3: sp.fungi:**
- 4: sp.human:**
- 5: sp.invertebrate:**
- 6: sp.mammal:**
- 7: sp.mhc:**
- 8: sp.organelle:**
- 9: sp.phage:**
- 10: sp.plant:**
- 11: sp.rodent:**
- 12: sp.virus:**
- 13: sp.vertibrate:**
- 14: sp.unclassified:**
- 15: sp.rvirus:**
- 16: sp.bacteriap:**
- 17: sp.archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	76.7	345	Q95LB0	Q95LB0 pan troglod
2	46	63.0	345	Q9CY42	Q9cy42 mus musculu
3	44	60.3	256	2_Q8GGJ1	Q8ggj1 lactobacill
4	44	60.3	785	3_Q9P8H5	Q9p8h5 humicola in
5	43	58.9	454	10_Q9S7B0	Q9s7b0 oryza sativ
6	42	57.5	574	5_Q19301	Q19301 caenorhabdi
7	42	57.5	1574	10_Q64923	Q64923 zea mays (m
8	41	56.2	519	10_Q8GZV1	Q8gzv1 oryza sativ
9	41	56.2	544	10_Q8H615	Q8h615 oryza sativ
10	41	56.2	1397	10_Q9LP90	Q9lp90 arabidopsis
11	40	54.8	359	2_Q8KH2	Q8khn2 clostridium
12	40	54.8	569	2_Q8KE55	Q8kr55 escherichia
13	40	54.8	719	2_Q9AEM4	Q9aem4 clostridium
14	40	54.8	1338	10_Q8RU03	Q8ruu3 oryza sativ
15	40	54.8	1750	10_Q9LPC5	Q9lpc5 arabidopsis
16	39	53.4	155	2_Q8KQ32	Q8kq32 pseudomonas

17	39	53.4	367	4	Q9H200	Q9h2q0 homo sapien
18	39	53.4	391	17	Q9HPW1	Q9hpx1 halobacteri
19	39	53.4	453	4	Q9P291	Q9p291 homo sapien
20	39	53.4	572	2	Q47908	Q47908 francisella
21	39	53.4	664	10	Q9ZQP2	Q9zqp2 arabidopsis
22	39	53.4	761	16	Q9F287	Q9f287 yersinia pe
23	39	53.4	1272	10	Q22731	Q22731 arabidopsis
24	38	52.1	25	5	Q9NBE2	Q9nbe2 chironomus
25	38	52.1	166	10	Q8R296	Q8rz96 oryza sativ
26	38	52.1	243	16	Q9FBK3	Q9fbk3 streptomyce
27	38	52.1	283	3	Q9UQX3	Q9uqx3 coriolus ve
28	38	52.1	372	17	Q8ZUR8	Q8zur8 pyrobaculum
29	38	52.1	488	16	Q98J30	Q98j30 rhizobium l
30	38	52.1	576	2	P75003	P75003 zoogloea ra
31	38	52.1	606	16	Q9WKN8	Q9wxn8 thermotoga
32	38	52.1	616	16	Q9I473	Q9i473 pseudomonas
33	38	52.1	1431	11	Q9JMH4	Q9jmh4 mesocricetu
34	38	52.1	1772	16	Q9X877	Q9x877 streptomyce
35	37	50.7	25	5	Q9NBE0	Q9nbe0 chironomus
36	37	50.7	98	5	Q9V1L9	Q9v1l9 chironomus
37	37	50.7	98	5	Q9V1L8	Q9v1l8 chironomus
38	37	50.7	98	5	Q9XTN1	Q9xtn1 chironomus
39	37	50.7	145	2	Q8VTC1	Q8vtc1 helicobacte
40	37	50.7	160	5	Q94443	Q94443 chironomus
41	37	50.7	160	6	Q95N39	Q95n39 saimiri sci
42	37	50.7	160	6	Q95N43	Q95n43 alouatta be
43	37	50.7	160	6	Q95N33	Q95n33 nycticebus
44	37	50.7	160	6	Q95N32	Q95n32 galago cras
45	37	50.7	160	6	Q95N42	Q95n42 ateles pani

ALIGNMENTS

RESULT 1

Q95LB0 PRELIMINARY; PRT; 345 AA.
AC Q95LB0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Apolipoprotein H.
GN APOH.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21372074; PubMed=11479737;
RA Sanghera D.K., Nestlerode C.S., Ferrell R.E., Kamboh M.I.;
RT "Chimpanzee apolipoprotein H (beta2-glycoprotein 1): report on the
gene structure, a common polymorphism, and a high prevalence of
antiphospholipid antibodies.";
RL Hum. Genet. 109:63-72 (2001).
DR EMBL; AF358415; AAK71538.1; JOINED.
DR EMBL; AF358408; AAK71538.1; JOINED.
DR EMBL; AF358409; AAK71538.1; JOINED.
DR EMBL; AF358410; AAK71538.1; JOINED.
DR EMBL; AF358411; AAK71538.1; JOINED.
DR EMBL; AF358412; AAK71538.1; JOINED.
DR EMBL; AF358413; AAK71538.1; JOINED.
DR EMBL; AF358414; AAK71538.1; JOINED.
DR InterPro; IPR000436; SUSHI_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
KW Lipoprotein.
SQ SEQUENCE 345 AA; 38267 MW; 712A3EDA2AD4FD36 CRC64;

Query Match 76.7%; Score 56; DB 6; Length 345;
Best Local Similarity 90.9%; Pred. No. 0.079;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 KKATGTHDG 11
Db 227 KKATGCHDG 237

RESULT 2
Q9CY42 PRELIMINARY; PRT; 345 AA.
AC Q9CY42;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 13 days embryo liver cDNA, RIKEN full-length enriched library,
DE clone:251008B09, full insert sequence.
GN APOH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Embryonic liver;
MEDLINE=21085660; PubMed=11217851;
Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Glass C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamaya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
Hayashizaki Y.,
"Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010926; BAB27271.1;
DR HSSP; P02749; 1C1Z.
DR MGD; MGI:88058; Apoh.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 4.
DR SMART; SM00032; CCP; 4.
SQ SEQUENCE 345 AA; 38688 MW; 33242A7E16FD99B CRC64;

Query Match 63.0%; Score 46; DB 11; Length 345;
Best Local Similarity 80.0%; Pred. No. 5;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKATGTHD 10
Db 227 KKATGCHGE 236

RESULT 3
Q8GGJ1 PRELIMINARY; PRT; 256 AA.
AC Q8GGJ1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Abc3.
DE Abc3.
GN ABC3.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCM3626;
RA Bringle F., Hubert J.-C.;
RT "Investigation of arginine requirement in Lactobacillus isolated from
RT different environments revealed point mutations, insertions and
RT deletions in arginine biosynthetic genes.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF514870; AAO15988.1;
SQ SEQUENCE 256 AA; 28597 MW; B4FE2B25276CEFF5C CRC64;

Query Match 60.3%; Score 44; DB 2; Length 256;
Best Local Similarity 57.1%; Pred. No. 8.2;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKATGTHDGGXA 14
Db 242 ODKLNFGAHGQVA 255

RESULT 4
Q9P8H5 PRELIMINARY; PRT; 785 AA.
AC Q9P8H5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cellobiose dehydrogenase.
OS Humicola insolens.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=34413;
[1]
SEQUENCE FROM N.A.
Gollightly E., Brown K.;
RT "Nucleic acids encoding polypeptides having cellobiose dehydrogenase
RT activity.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Xu F., Gollightly E.J., Duke K.R., Lassen S.F., Brown K.M., Brown S.H.,
RA Shulein M.;
RT "Humicola insolens cellobiose dehydrogenase: cloning, redox chemistry,
RT and 'logic gate'-like dual functionality.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF257654; AAF69005.1;
DR HSSP; Q01738; 1D7C.
DR InterPro; IPR000172; GMC_oxred.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00732; GMC_oxred; 1.
DR PROSITE; PS00623; GMC_OXRED_1; 1.
SQ SEQUENCE 785 AA; 84856 MW; 223F5B89A122FEAF CRC64;

Query Match 60.3%; Score 44; DB 3; Length 785;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 DKATGTHDGG 12
Db 188 DTATGFDHNG 198

RESULT 5
Q9S7B0 PRELIMINARY; PRT; 454 AA.
AC Q9S7B0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.

```

OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0711E10."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0705D01."
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP000570; BA85210.1;
 DR EMBL; AP000492; BA84635.1;
 DR Gramene; Q9S7B0;
 SQ SEQUENCE 454 AA; 47854 MW; B56214B992CD5C88 CRC64;

 Query Match 58.98; Score 43; DB 10; Length 454;
 Best Local Similarity 80.08; Pred. No. 24;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 QY 5 TFGTHDGGXA 14
 | | | | | | | |
 Db 127 TEGTHDGGAA 136

 RESULT 6
 Q19301
 ID Q19301 PRELIMINARY; PRT; 574 AA.
 AC Q19301;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical 62.3 kDa protein.
 GN F10E7.9.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RX MEDLINE-99069613; PubMed-9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Pauley A.;
 RT "The sequence of C. elegans cosmid F10E7.";
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U41264; AA82423.1;
 DR WormPep; F10E7.9; CE04364.
 DR InterPro; IPR002293; AA/rel_permeasel.
 KW Hypothetical protein.
 SQ SEQUENCE 574 AA; 62301 MW; AA209D1029FC700E CRC64;

 Query Match 57.58; Score 42; DB 5; Length 574;
 Best Local Similarity 72.78; Pred. No. 46;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

 QY 1 KDKATFGTHDG 11

Db 42 KDKVLEGTWDG 52
 ||| ||| ||

 RESULT 7
 Q64923
 ID Q64923 PRELIMINARY; PRT; 1674 AA.
 AC Q64923;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Starch synthase DULL1.
 GN DULL1.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98169346; PubMed-9501113;
 RA Gao M., Wanat J., Stinaard P.S., James M.G., Myers A.M.;
 RT "Characterization of dull1, a maize gene coding for a novel starch
 RT synthase.";
 RL Plant Cell 10:399-412(1998).
 DR EMBL; AF023159; AAC14014.1;
 DR InterPro; IPR001296; glyco_trans_1.
 DR Pfam; PF00534; Glycos_transf_1; I.
 SQ SEQUENCE 1674 AA; 188311 MW; FA31BE959A01771D CRC64;

 Query Match 57.58; Score 42; DB 10; Length 1674;
 Best Local Similarity 72.78; Pred. No. 1.5e+02;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

 QY 1 KDKATFGTHDG 11
 | | | | | | | |
 Db 611 KQLATVGTGHDG 621

 RESULT 8
 Q8GZY1
 ID Q8GZY1 PRELIMINARY; PRT; 519 AA.
 AC Q8GZY1;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN OSJNBA0090010.1.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Currie J.,
 RA Collura K.;
 RT "Rice genomic Sequence.";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC134237; AA017011.1;
 KW Hypothetical protein.
 SQ SEQUENCE 519 AA; 56443 MW; 6F065D65D838D013 CRC64;

 Query Match 56.28; Score 41; DB 10; Length 519;
 Best Local Similarity 70.08; Pred. No. 63;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

 QY 3 KATFGTHDGG 12
 : | | | | | | | |
 Db 192 RAAFGRDGG 201

 RESULT 9

Q8H615 Q8H615 PRELIMINARY; PRT; 544 AA.
 AC Q8H615;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE OSJNBA0035103.11 protein.
 GN OSJNBA0035103.11.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GAS) genomic DNA, chromosome 6, BAC
 RT clone:OSJNBA0035103.",
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003019; BAC22238.1; -;
 SQ SEQUENCE 544 AA; 58723 MW; ECDE695F0E22A269 CRC64;

Query Match 56.2%; Score 41; DB 10; Length 544;
 Best Local Similarity 58.3%; Pred. No. 66;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDKATFGTHDGG 12
 Db 117 QDKLVGTGDDG 128

RESULT 10
 Q9LP90
 ID Q9LP90 PRELIMINARY; PRT; 1397 AA.
 AC Q9LP90;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 22, Last annotation update)
 DE T32E20.30.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Brassicales; Brassicaceae; Arabidopsia.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
 RA Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
 RA Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC T32E20 from chromosome
 RT 1.",
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]

SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,

RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC020646; AAF79797.1; -;
 DR InterPro; IPR000953; Chromo.
 DR InterPro; IPR005162; Retrotrans_gag.
 DR InterPro; IPR001584; Rve.
 DR InterPro; IPR000477; RVTse.
 DR Pfam; PF00385; chromo; 1.
 DR Pfam; PF03732; Retrotrans_gag; 1.
 DR Pfam; PF00665; rve; 1.
 DR Pfam; PF00078; rvt; 1.
 DR SMART; SM00298; CHROMO; 1.
 DR PROSITE; PS50013; CHROMO_2; 1.
 KW RNA-directed DNA polymerase; Transferase.
 SQ SEQUENCE 1397 AA; 159426 MW; 1ED5717280BF298E CRC64;

Query Match 56.2%; Score 41; DB 10; Length 1397;
 Best Local Similarity 70.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 DKATFGTHDG 11
 Db 634 EXTFTTHDG 643

RESULT 11
 Q8KH2
 ID Q8KH2 PRELIMINARY; PRT; 359 AA.
 AC Q8KH2;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE S-layer protein (Fragment).
 GN SLPA.
 OS Clostridium difficile.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1496;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 43596, and 630;
 RX MEDLINE=22083941; PubMed=12089261;
 RA Karjalainen T., Saumier N., Barc M.C., Delmee M., Collignon A.;
 RT "Clostridium difficile genotyping Based on alpa Variable Region in S-
 RT Layer Gene Sequence: an Alternative to Serotyping".
 RL J. Clin. Microbiol. 40:2452-2458(2002).
 DR EMBL; AF448123; AAM75923.1; -;
 DR EMBL; AF448124; AAM75924.1; -;
 FT NON_TER 1
 FT NON_TER 359
 SQ SEQUENCE 359 AA; 38298 MW; 8A72107B4752DB4B CRC64;

Query Match 54.8%; Score 40; DB 2; Length 359;
 Best Local Similarity 66.7%; Pred. No. 63;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDKATFGTHDGG 12
 Db 153 KDVATFGLKSGG 164

RESULT 12
 Q8KR55
 ID Q8KR55 PRELIMINARY; PRT; 569 AA.
 AC Q8KR55;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Iucc.
 GN Iucc.
 OS Escherichia fergusonii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.

```

OX NCBI_TaxID=564;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EF873;
RA Smajd D., Smarda J., Weinstein G.M.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046057; AAL01548.1; -
SQ SEQUENCE 569 AA; 64344 MW; 78E6660D08E169CA CRC64;

Query Match 54.8%; Score 40; DB 2; Length 569;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 KATFGTHDGG 12
DB 536 KLTFSHDGG 545

RESULT 13
Q9AEM4 PRELIMINARY; PRT; 719 AA.
ID Q9AEM4
AC Q9AEM4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE S-layer precursor protein.
GN SLPA.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C253;
RX MEDLINE=21189268; PubMed=11292772;
RA Karjalainen T.K., Waligora-Dupriet A.J., Cerquetti M., Spigaglia P.,
RA Mauri P., Mastrantonio P.;
RT "Molecular and genomic analysis of two genes encoding surface-anchored
RT proteins from clostridium difficile.";
RL Infect. Immun. 69:3442-3446(2001).
DR EMBL: AJ291709; CAC36294.1; -
FT S-layer; Signal.
FT SIGNAL 1 24 POTENTIAL.
SQ SEQUENCE 719 AA; 76133 MW; DBE67B738C2F8A20 CRC64;

Query Match 54.8%; Score 40; DB 2; Length 719;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 KDKATFGTHDGG 12
DB 171 KDVATFGLKSG 182

RESULT 14
Q8RUU3 PRELIMINARY; PRT; 1338 AA.
ID Q8RUU3
AC Q8RUU3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Putative gag-pol polyprotein.
GN QJ1004.F02.14 OR OSJNBA0014J14.7.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,

```

```

RA Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pal G.,
RA Vanaken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OJ1004.F02 genomic sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Wang R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Saski C., Henry D., Oates R., Simmons J.;
RT "Rice Genomic Sequence.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC092387; AAL82656.1; -
DR EMBL: AC092172; AAM18147.1; -
DR Gramene; Q8RUU3; -
DR InterPro; IPR000953; Chromo.
DR InterPro; IPR005162; Retrotrans_gag.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF03732; Retrotrans_gag; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00078; rvt; 1.
DR SMART; SM00298; CHROMO; 1.
KW Polyprotein; RNA-directed DNA polymerase; Transferase.
SQ SEQUENCE 1338 AA; 150643 MW; 7D68E2A564279648 CRC64;

Query Match 54.8%; Score 40; DB 10; Length 1338;
Best Local Similarity 70.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 DKATFGTHDG 11
DB 513 DKTAERTHDG 522

RESULT 15
Q9LPC5 PRELIMINARY; PRT; 1750 AA.
ID Q9LPC5
AC Q9LPC5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE F22M8.9 protein.
GN F22M8.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Liu S.X., Sakano H., Yu G., Lee J.M., Lenz C., Pham P., Toriumi M.,
RA Chin C., Chlou J., Choi E., Chung M., Gonzalez A., Hwang B., Liu A.,
RA Vaysberg M., Alfalfi H., Brooks S., Buehler E., Chao Q., Conn L.,
RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "The sequence of BAC F22M8 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC020622; AAF76474.1; -
DR HSP; Q99418; 1PBV.
DR InterPro; IPR006162; Ppantne_attach.
DR InterPro; IPR000904; Sec7.
DR Pfam; PF01369; Sec7; 1.
DR SMART; SM00222; Sec7; 1.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR PROSITE; PS50190; SEC7; 1.
SQ SEQUENCE 1750 AA; 194941 MW; F4FCFF663ACDB019 CRC64;

Query Match 54.8%; Score 40; DB 10; Length 1750;
Best Local Similarity 77.8%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 3 KATFGHDG 11
| | | | |
Db 264 KTFGGHDG 272

Search completed: August 28, 2003, 18:38:00
Job time : 34.7273 secs